UDC 575.630
https://doi.org/10.2298/GENSR1702613S
*Original scientific paper*

# GENETIC STRUCTURE AND DIVERSITY OF EUROPEAN CHESTNUT (*Castanea sativa* MILL.) POPULATIONS IN WESTERN BALKANS: ON A CROSSROAD BETWEEN EAST AND WEST

Azra SKENDER [1], Mirsad KURTOVIĆ [2], Naris POJSKIĆ [3], Belma KALAMUJIĆ STROIL [3*], Semina HADŽIABULIĆ[4], Fuad GAŠI[2]

[1]Biotechnical Faculty, University of Bihać
[2]Faculty of Agriculture and Food Sciences, University of Sarajevo
[3]Institute for Genetic Engineering and Biotechnology, Laboratory for Molecular Genetics of Natural Resources, University of Sarajevo
[4]Agromediterranean Faculty, University Džemal Bijedić of Mostar

Skender A., M. Kurtović , N. Pojskić, B. Kalamujić Stroil, S. Hadžiabulić, F. Gaši (2017): *Genetic structure and diversity of European chestnut (Castanea sativa mill.) populations in Western balkans: on a crossroad between east and west*.- Genetika vol49, no2, 613- 626.

European chestnut (*Castanea sativa* Mill.) is highly valued in the western Balkans as a source of timber and fruit, but also as an important source of nectar and pollen for the production of honey. In this study, four chestnut populations, covering a major portion of the western Balkans, and a reference population from the northern Italy were examined using 21 microsatellite markers. The highest genetic diversity was detected within the populations geographically closest to the Italian Peninsula, which also displayed the highest level of admixture with the samples from Italy. The strongest genetic differentiation was noted among the southern and eastern chestnut populations from Bosnia and Herzegovina (B&H) (Gst=12.05%). This pronounced differentiation is probably caused by the genetic adaptations to notably different climatic conditions present in the south (Mediterranean climate) and east (Continental climate) of B&H. The clear genetic differentiation of the southern and eastern B&H chestnut populations from the Italian population, determined by pairwise Gst, FCA and Bayesian Structure analyses, indicates that these populations most likely originated from independent shelter zones (refugia), after the last glaciation period. Based on these results we propose a presence of an introgression zone in the northwestern Balkans, established through gene flow from the Italian and the Balkan Peninsula. The obtained insights into the structure

*Corresponding author:* Belma Kalamujić Stroil, Institute for Genetic Engineering and Biotechnology, University of Sarajevo, Zmaja od Bosne 8, 71000 Sarajevo, Bosnia and Herzegovina, Phone:+38733220926, fax:+38733442891,E-mail: belma.kalamujic@ingeb.unsa.ba

of all analyzed populations will significantly contribute towards establishing a regional conservation and utilization strategy for European chestnut in western Balkans.

*Keywords*: genetic differentiation, European chestnut, introgression zone, microsatellites, refugia

## INTRODUCTION

European chestnut (*Castanea sativa* Mill.) is a tree species, widely present throughout all the Mediterranean peninsulas. The Quaternary cold periods, which reduced the woodland diversity in Europe to restricted number of shelter zones (refugia), located in the south of the Continent (ORDONEZ and SVENNING, 2017), significantly influenced the current structure and distribution of chestnut populations. Aside from the natural post-glacial expansion from these refugia, the cultivation of *C. sativa*, especially during medieval times (CONEDERA *et al*., 2004), may have also impacted the distribution of this species. The main chestnut refugia, identified using an extended palynological approach by KREBS *et al*. (2004), are located in the Transcaucasian region, as well as on the Apennine and the Iberian peninsulas. The same study also reported a high likelihood of *Castanea* refugia on the Balkan Peninsula. Molecular data obtained from the microsatellite study of 31 *C. sativa* populations, positioned throughout southern Europe and Turkey, largely confirmed the previously hypothesized shelter zones (MATTIONI *et al*., 2013). The aforementioned molecular study also reported that the chestnut populations from western Turkey and southern Balkans (Greece) clustered tightly together, while the Italian and Iberian Peninsula populations formed a single, distinct cluster. The observed differentiation between the western and the eastern chestnut populations ruled out, both the natural and the human-mediated transfer of germplasm from Greek areas to the Italian Peninsula. However, both the palynological and the molecular study cited here did not include samples from western Balkans, an area which geographically links Italy to the rest of the Balkan Peninsula.

The largest concentration of *C. sativa*, in the western part of the Balkan Peninsula, is located on the border of Slovenia and Croatia, around the river Sava (Figure 1). More sporadic and somewhat isolated populations are also present in several parts of Bosnia and Herzegovina (B&H). In both of these regions, chestnut is highly valued as a source of timber and fruit, but also as an important source of nectar and pollen for the production of honey (IDŽOJTIĆ *et al*., 2009). During the 1970s several promising *C. sativa* types, selected from the natural populations in western Balkans, were grafted and utilized in nursery production. Still, the area currently occupied by chestnut orchard in western Balkans is negligible in comparison to the range of existing natural populations. In terms of the overall production, the countries located in the western Balkan Peninsula fall far behind Spain, Portugal, Italy and Turkey, the world leading producers of European chestnut. Nonetheless, due to the geographic location of the Balkans, on a crossroad between east and west, *C. sativa* populations found here could potentially represent an interesting source of chestnut genetic diversity.

The traditional approach to examining chestnut population, based on morphologic observation, has largely been supplemented or substituted with the use of molecular markers. In recent years, microsatellite markers have been the most frequently used tool for assessing the diversity of chestnut populations. Microsatellites or SSRs (simple sequence repeats) have previously proved valuable for estimating the genetic diversity and structure of *C. sativa* populations in Spain, Italy, Switzerland, Greece, Bulgaria and Turkey (MARTÍN *et al*., 2010; BECCARO *et al*., 2012; MARTÍN *et al*., 2012; MATTIONI *et al*., 2013; LUSINI *et al*., 2014). The same

markers have also been used for the genetic characterization of grafted chestnut accessions from Portugal, Spain and Italy (PEREIRA-LORENZO *et al*., 2010; MARINONI *et al*., 2013). Microsatellites chosen for studies on chestnut are dominantly selected from two sets of markers, developed in *Castanea sativa* by MARINONI *et al*. (2003) and BUCK *et al*. (2003). However, until now, only a few of these SSRs have been applied in large-scale studies on chestnut populations.

In this study we used a set of 21 microsatellite markers, in order to: 1) assess the genetic diversity and structure of European chestnut populations in western Balkans, 2) investigate the relationship between a northern Italian and western Balkans *C. sativa* populations, with the purpose of shedding additional light on the expansion process of European chestnut after the last glaciation period.

## MATERIALS AND METHODS

### *SSR analyses*

Total of 171 *Castanea sativa* individuals from five chestnut populations (Table 1) were used for the molecular analyses. Four of the five examined populations are spread throughout the western Balkans, while the fifth, used as a reference population, is located in northern Italy. Plant material (leaves) for the DNA extraction was collected from randomly selected plants on each analyzed site. Each sampled tree, within each site, was at least 20 m apart from all other sampled trees. Leaves were harvested during spring and summer and pulverized using mortar and pestle. DNA was isolated from approximately 50 mg of leaf powder according to DOYLE and DOYLE (1987) CTAB protocol. Twenty one SSR markers, described in MARINONI *et al*. (2003) (CsCAT set) and BUCK *et al*. (2003) (EMC set), were chosen based on their polymorphism reported in previous studies.

*Table 1. Number of sampled trees and geographical location of five European chestnut populations investigated in this study.*

| *Country* | *Population / collection site* | *Code* | *No. of sampled trees* |
|---|---|---|---|
| *Bosnia and Herzegovina* | *Pećigrad* | *CP* | *37* |
| *Bosnia and Herzegovina* | *Konjic* | *CK* | *45* |
| *Bosnia and Herzegovina* | *Bratunac* | *CB* | *39* |
| *Croatia and Slovenia* | *Sava valley* | *SC* | *30* |
| *Italy (South Tyrol)* | *Merano* | *IT* | *20* |

All PCR reactions were conducted in total volume of 15 µl, containing 2 mM of $MgCl_2$, 1 x PCR buffer, 0.2 mM of dNTPs, 0.05 U/µl of TrueStart *Taq* polymerase (Thermo Scientific) and 10-50 ng of template DNA. CsCAT set of primer pairs was amplified according to protocol in Marinoni et al. (2003), with minor modifications. Thermal conditions for amplification of EMC loci were as described in Buck et al. (2003). All PCR reactions were carried out in GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA, USA). Aliquots of amplified DNA (1 µl), mixed with formamide and GeneScan-350 ROX Size Standard (Applied Biosystems) were run on an ABI Prism 310 automated sequencer (Applied Biosystems). SSR profiles were scored using the GeneMapper Software ID v3.2 (Applied Biosystems). In case of any uncertainty regarding the scoring process, PCR amplification was repeated using only one primer pair per reaction.

***Data analyses***

Number of alleles per locus (Na), effective number of alleles per locus (Ne), observed (Ho) and expected (He) heterozygosity, unbiased expected (UHe) heterozygosity and inbreeding coefficient (Fis) were calculated using the program GenAlEx 6.5 (PEAKALL and SMOUSE, 2012), while the allelic richness (Rs) was assessed using FSTAT (GOUDET, 2002). Fisher's exact test by means of the Markov Chain algorithm (GUO and THOMPSON, 1992) was used to estimate deviations from the Hardy-Weinberg equilibrium.

Presence of null alleles, long allele's dropout or scoring errors due to stuttering in the SSRs profile were tested using Micro-checker 2.2.3 (VAN OOSTERHOUT *et al*., 2004). Correction of allele frequencies across loci due to null allele presence was done with ML-NullFreq (KALINOWSKI and TAPER, 2006). The procedure includes maximum likelihood estimation of the frequency of null alleles at microsatellite loci.

Population differentiation across loci (Gst) and pairwise Gst were calculated according to Nei (1987) using GenAlEx 6.5 (PEAKALL and SMOUSE, 2012) and Dispan (OTA, 1993) software. Mantel test (MANTEL, 1967) with 5,000 permutations, implemented within PAST ver. 2.17.c. (HAMMER *et al*., 2001), was used in order to estimate compatibility of pairwise Gst matrix before and after allele frequencies correction due to null allele presence.

A multivariate analysis, FCA (Factorial Correspondence Analysis) based on allele frequencies, was performed using Genetix 4.02 (BELKHIR *et al*., 2001).

In order to assign individual genotypes to a specific population, the Bayesian model-based cluster procedure within Structure version 2.2.3 (PRITCHARD *et al*., 2000) was used. We computed K (unknown) clusters of individuals testing K (log-likelihood) = 1–10 for all accessions, assuming that sampled trees were of unknown origin. Assignment of individuals to a cluster was provided by a probability of membership *qI* chosen at 75%, as proposed by MATTIONI *et al*. (2013). Tests were done based on an admixture model where the allelic frequencies were correlated and a burn-in period of 200,000 and 500,000 iterations for data collection was applied. In order to estimate the most probable K value for the analyzed data we used the method described by EVANNO *et al*. (2005), implemented through Structure Harvester ver. 0.6. application (EARL and VON HOLDT, 2011). A hierarchical AMOVA (EXOFFIER *et al*., 1992) was carried out, implemented with Arlequin ver. 3.5.1.2 (EXCOFFIER and LISCHER, 2010), based on different partitioning of variance. The analyses of molecular variance included all individuals with the probability of membership to a cluster (K=3) above 75%.

## RESULTS

***SSR polymorphism***

Twenty one microsatellite loci, analyzed in this study, greatly varied in regards to allele diversity (Table 2). The highest total number of alleles per locus was detected for CsCAT3 (13), while the lowest number of different alleles was registered for EMCs11 (2). Overall number of alleles, obtained using 21 primer pairs on 171 individual chestnut trees, was 125. Allele frequencies for all analyzed loci, per population, are given in Online Resource 1. Mean gene diversity, calculated over all loci, was moderately high (0.605) and ranged from 0.081 for EMCs14 to 0.814 for EMCs38. The overall value of fixation index (Fis), which represents the measure of heterozygote deficit, was positive (0.277), whereas only two loci displayed negative values for this measure (EMCs14 and CsCAT17).

*Table 2. Description of diversity for each of the 21 analyzed microsatellite loci. Number of alleles per locus (Na), effective number of alleles per locus (Ne), observed (Ho) and expected (He) heterozygosity, unbiased expected (UHe) heterozygosity and inbreeding coefficient (Fis).*

| *Locus* | *Na* | *Ne* | *Ho* | *He* | *UHe* | *Fis* |
|---|---|---|---|---|---|---|
| EMCs2[ab] | 3 | 2.27 | 0.444 | 0.560 | 0.562 | 0.207 |
| EMCs4 | 4 | 2.26 | 0.526 | 0.558 | 0.559 | 0.056 |
| EMCs10 | 5 | 2.24 | 0.433 | 0.555 | 0.556 | 0.220 |
| EMCs11[a] | 2 | 1.92 | 0.337 | 0.480 | 0.481 | 0.297 |
| EMCs13 | 3 | 1.57 | 0.205 | 0.362 | 0.363 | 0.435 |
| EMCs14[a] | 3 | 1.09 | 0.084 | 0.081 | 0.081 | -0.033 |
| EMCs15[b] | 4 | 2.47 | 0.474 | 0.595 | 0.597 | 0.204 |
| EMCs17[b] | 5 | 3.21 | 0.673 | 0.689 | 0.691 | 0.023 |
| EMCs22[ab] | 8 | 2.98 | 0.471 | 0.664 | 0.666 | 0.290 |
| EMCs25 | 5 | 3.82 | 0.143 | 0.738 | 0.741 | 0.807 |
| EMCs32[a] | 7 | 2.58 | 0.378 | 0.612 | 0.615 | 0.382 |
| EMCs38 | 9 | 5.36 | 0.360 | 0.814 | 0.816 | 0.558 |
| EMCs42[b] | 3 | 2.10 | 0.361 | 0.524 | 0.525 | 0.310 |
| CsCAT1[a] | 9 | 3.50 | 0.420 | 0.714 | 0.716 | 0.412 |
| CsCAT2 | 9 | 4.11 | 0.714 | 0.756 | 0.759 | 0.056 |
| CsCAT3[a] | 13 | 5.02 | 0.292 | 0.801 | 0.804 | 0.635 |
| CsCAT4 | 4 | 2.32 | 0.423 | 0.569 | 0.570 | 0.255 |
| CsCAT14[a] | 7 | 2.86 | 0.437 | 0.650 | 0.652 | 0.327 |
| CsCAT15[a] | 6 | 2.38 | 0.436 | 0.580 | 0.582 | 0.249 |
| CsCAT16[a] | 6 | 2.88 | 0.529 | 0.653 | 0.655 | 0.190 |
| CsCAT17[a] | 10 | 4.09 | 0.796 | 0.755 | 0.757 | -0.055 |
| Mean | 5.95 | 2.91 | 0.426 | 0.605 | 0.607 | 0.277 |

[a]SSR loci mapped on the same linkage group mapped by Barreneche et al. (2004) and Casasoli et al. (2006) as at least one other of the 21 loci analyzed in this study.

[b]SRR loci not deviating from Hardy-Weinberg equilibrium in any of the analyzed populations and not affected by the null alleles.

The set of microsatellites isolated by MARINONI *et al*. (2003) (CsCAT), yielded on average 8 alleles per locus, while the set isolated by BUCK *et al.* (2003) (EMCs) was considerably less polymorphic (on average 4.7 alleles per locus). Similar situation was observed for gene diversity, which was lower for EMCs loci (0.556) than for the CsCAT loci (0.685). However, all eight CsCAT markers, analyzed in this study, represent dinucleotide microsatellites, while seven of the twelve examined EMCs SSRs contain a trinucleotide repeat motif (EMCs 17 contained a compound microsatellite of a trinucleotide and a tetranucleotide repeat). In order to conduct a fair

assessment of CsCAT and EMCs microsatellites, trinucleotide microsatellites, which mutate at a lower rate than dinucleotide (CHAKRABORTY *et al*., 1997), were excluded from the comparison. When only EMCs loci with dinucleotide repeats were examined, the value for the average number of alleles per locus was somewhat higher (6.2 alleles per locus), as was the gene diversity (0.651).

Only five of the 21 SSR loci, analyzed in this study, did not deviate from Hardy-Weinberg equilibrium (HWE) in any of the examined populations. Micro-checker analyses identified the presence of null alleles in each of the loci deviating from HWE, but did not detect allele dropout or scoring errors. The effect of the presence of null alleles, detected among most of the analyzed loci, was taken into the consideration for the estimates of genetic differentiation and population structure.

***Population diversity and differentiation***

Genetic diversity parameters calculated for each of the five analyzed chestnut populations are presented in the Table 3. The highest number of alleles per locus was found in the Sava valley population (SC) (4.76), which is located on the boarder of Slovenia and Croatia, while the lowest number was detected in the Bratunac population (CB) (3.86), located in the eastern part of Bosnia and Herzegovina. In regards to the effective number of alleles per locus and the allelic richness, populations from Croatia-Slovenia and Italy (SC and IT) displayed higher values than the three populations from Bosnia and Herzegovina. The only exception was the northwestern B&H population of Pećigrad (CP), which is in fact geographically much closer to the Sava valley population than to the other two sampled sites in Bosnia and Herzegovina (Figure 1). In regards to the observed and expected heterozygosity, as well as the unbiased expected heterozygosity, of all three B&H populations, only Pećigrad registered similarly high values for these parameters as the ones calculated for the chestnut from Croatia-Slovenia and Italy. The two non-B&H populations also displayed lower levels of heterozygosity deficiency (low Fis values). Comparable values for the fixation index were obtained analyzing just the five loci not deviating from the Hardy-Weinberg equilibrium. The decrease in heterozygosity noted among the southern (CK) and eastern (CB) chestnut populations in Bosnia and Herzegovina indicates that these populations are genetically more isolated than the northwestern one (CP).

*Table 3. Number of alleles per locus (Na), effective number of alleles per locus (Ne), allelic richness (Rs), observed (Ho) and expected (He) heterozygosity, unbiased expected (UHe) heterozygosity and inbreeding coefficient (Fis), calculated over 21 microsatellite loci.*

| *Population* | *Country* | *Na* | *Ne* | *Rs* | *Ho* | *He* | *uHe* | *Fis*[a] | *Fis*[b] |
|---|---|---|---|---|---|---|---|---|---|
| Pećigrad (CP) | Bosnia and Herzegovina | 4.48 | 2.75 | 3.79 | 0.434 | 0.580 | 0.588 | 0.246 | 0.045 |
| Konjic (CK) | Bosnia and Herzegovina | 4.67 | 2.15 | 3.56 | 0.368 | 0.454 | 0.459 | 0.172 | 0.003 |
| Bratunac (CB) | Bosnia and Herzegovina | 3.86 | 2.28 | 3.20 | 0.391 | 0.473 | 0.480 | 0.148 | 0.079 |
| Sava valley (SC) | Croatia and Slovenia | 4.76 | 2.49 | 4.10 | 0.493 | 0.546 | 0.557 | 0.103 | -0.069 |
| Merano (IT) | Italy | 4.62 | 3.01 | 4.34 | 0.530 | 0.594 | 0.613 | 0.084 | -0.006 |

[a]Fis calculated over all 21 SSR loci included in this study.
[b]Fis calculated over five SSR loci not deviating from Hardy-Weinberg equilibrium and without the presence of null alleles.

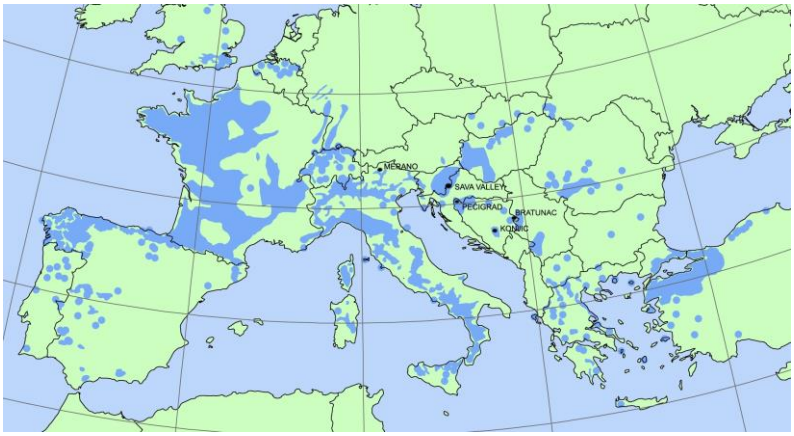


Figure 1. Geographical distribution of the five *Castanea sativa* populations examined in the study. The collection site of each population is marked on the map. The distribution map was downloaded from the EUROPEAN FOREST GENETIC RESOURCES PROGRAMME website (http://www.euforgen.org/distribution_maps.html).

Eleven of the 21 SSR loci, examined here, have previously been mapped by BARRENECHE *et al*. (2004) and CASASOLI *et al*. (2006) on the same chestnut linkage group as at least one other microsatellite loci included in this study. Therefore, the data obtained from six of these eleven loci (EMCs2, EMCs11, EMCs22, EMCs32, CsCAT14 and CsCAT15) was removed for the estimates of genetic differentiation between the chestnut populations. Selecting which one of the linked microsatellites should be discarded was based on their polymorphism. Consequently those markers with the highest number of alleles detected in our study remained for further analyses.

Pairwise Gst, calculated over 15 remaining SSR loci, provided insight into the genetic differentiation among each pair of the five analyzed chestnut populations (Table 4). The strongest differentiation, based on this parameter, was detected among the southern and eastern populations from Bosnia and Herzegovina, Konjic and Bratunac, while a much weaker differentiation was registered among the populations from Croatia-Slovenia and Italy. However, the lowest Gst value was observed between the Italian population and the northwestern B&H population (Pećigrad).

*Table 4. Pairwise Gst calculated on five analyzed chestnut populations, over 15 unlinked SSR loci, before (upper triangle) and after (lower triangle) the correction for the presence of null alleles using ML-NullFreq (Kalinowski and Taper 2006).*

| | Pećigrad (CP) | Konjic (CK) | Bratunac (CB) | Sava valley (SC) | Merano (IT) |
|---|---|---|---|---|---|
| Pećigrad (CP) | | 0.0754 | 0.0696 | 0.0487 | 0.0228 |
| Konjic (CK) | 0.0713 | | 0.1435 | 0.0915 | 0.0729 |
| Bratunac (CB) | 0.0650 | 0.1205 | | 0.0780 | 0.0704 |
| Sava valley (SC) | 0.0538 | 0.0786 | 0.0634 | | 0.0423 |
| Merano (IT) | 0.0327 | 0.0684 | 0.0685 | 0.0509 | |

The previously conducted Micro-checker analyses identified null alleles among twelve of the 15 selected SSR loci. The presence of null alleles can lead to overestimation of genetic differentiation among groups and populations, especially in cases of low levels of gene flow (CHAPUIS and ESTOUP, 2007). Therefore a correction was applied for their presence using ML-NullFreq (KALINOWSKI and TAPER, 2006). The corrected allele frequencies were subsequently used for calculating pairwise Gst. By comparing the pairwise Gst values, calculated on the original and the corrected SSR data, the largest overestimation caused by null alleles was identified among two populations from Bosnia and Herzegovina, Konjic and Bratunac (Table 4). These two populations also displayed the strongest genetic differentiation among all five analyzed in this study. On the other hand, the genetic differentiation among populations from Croatia-Slovenia, Italy and northwestern B&H (Pećigrad), which was comparably low, was even slightly underestimated by the presence of null alleles, indicating an intense gene flow between these sites. Despite the mentioned discrepancies in differentiation estimates, a Mantel test showed a very strong correlation ($r=0.9875$, $P<0.01$) between pairwise Gst matrices before and after allele frequencies correction due to null allele presence. The overall genetic differentiation (total Gst), among all five of the analyzed populations, based on the uncorrected SSR data (12.39%) was higher than the differentiation calculated over allele frequencies corrected for the presence of null alleles (10.29%). Thus, the obtained results confirm the exaggerating effect of null alleles on the overall genetic differentiation.

***Population structure***

SSR data, obtained after correcting for the presence of null alleles with ML-NullFreq (KALINOWSKI and TAPER, 2006), was comprised of allele frequencies. The available, corrected allele frequencies are suitable for conducting the estimates of genetic differentiation, since the bias caused by null alleles has been excluded. However, statistical methods, such as Bayesian structure analyses and factorial correspondence analyses, which are commonly used in studies on chestnut germplasm (PEREIRA-LORENZO *et al*., 2010; BECCARO *et al*., 2012; MARTÍN *et al*., 2012; MATTIONI *et al*., 2013), require genotype data and not allele frequencies as input. Therefore, in order to further investigate the structure of the analyzed chestnut populations, while excluding any possible bias, only the genotype data obtained from the following five SSR loci was used: EMCs2, EMCs15, EMCs17, EMCs 22 and EMCs42. These five SSR loci were not affected by null alleles, they do not deviate from the Hardy-Weinberg equilibrium, and they have not previously been mapped on the same linkage group.

With the aim of gaining insight into the genetic relationships between the five analyzed chestnut populations, a factorial correspondence analyses (FCA) was performed on the data from the five selected SSR loci (Figure 2). The chestnut populations from Croatia-Slovenia and Italy generally overlapped with each other, and also with the northwestern B&H population Pećigrad. In contrast, a clear separation could be noted between the southern (Konjic) and eastern (Bratunac) B&H populations. This is entirely in accordance with the results of pairwise Gst analysis, calculated over 15 SSR loci, which detected the strongest genetic differentiation between Konjic and Bratunac (Gst=12.05%), and the weakest among the samples from Sava valley (Croatia-Slovenia), north Italy and Pećigrad (Gst=5.09% and 3.27% respectively). The FCA also demonstrated that most of the samples from Konjic differentiated clearly from all other populations analyzed in this study.

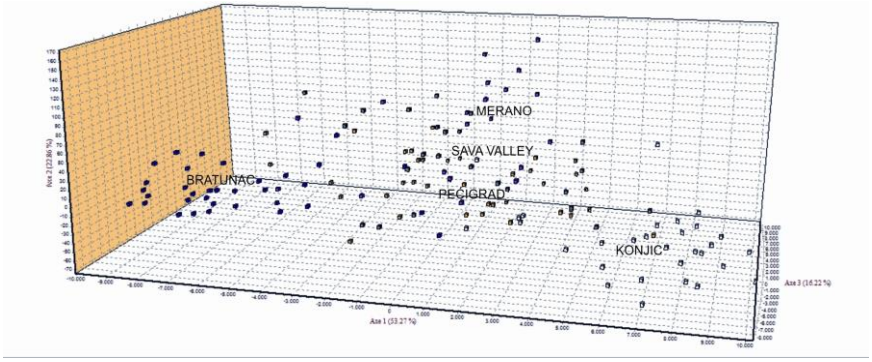


Figure 2. Multivariate analysis (Factorial Correspondence Analysis) of SSR data for 171 European chestnut samples, collected from five *Castanea sativa* populations examined in the study. The name of the populations is inserted next to the data point that signifies the center of each analyzed population.

Bayesian analysis, done within Structure, was also based on five SSR loci not deviating from HWE and included five populations with a total of 171 chestnut samples. Subsequent ΔK analyses (EVANNO *et al*., 2005) revealed maximum values on K=3. The southern and eastern B&H populations, Konjic and Bratunac, dominantly grouped into two distinct clusters, while the samples from the Italian population grouped inside a third cluster (Figure 3). Both Konjic and Bratunac populations have displayed equal genetic differentiation from the Italian population (Gst=6.84% and 6.85%, respectively) in the previous pairwise Gst analyses. The northwestern B&H chestnut population, Pećigrad and the population from Croatia-Slovenia adhered mainly to the third, but also to a lesser extent to the second cluster, showing a higher degree of admixture. Results obtained from the Bayesian analysis confirmed the presence of strong gene flow among populations from Croatia-Slovenia, Italy and northwestern B&H (Pećigrad), previously indicated by low pairwise Gst values. Konjic was clearly identified as the most differentiated population, while the introgression of the Bratunac gene pool was noted in the Sava valley and Pećigrad population. The strong differentiation among Structure clusters was also confirmed by AMOVA ($f_{CT}$=0.223; P<0.01) (Table 5). The analyses of molecular variance included all individuals with the probability of membership to a cluster (K=3) above 75%.

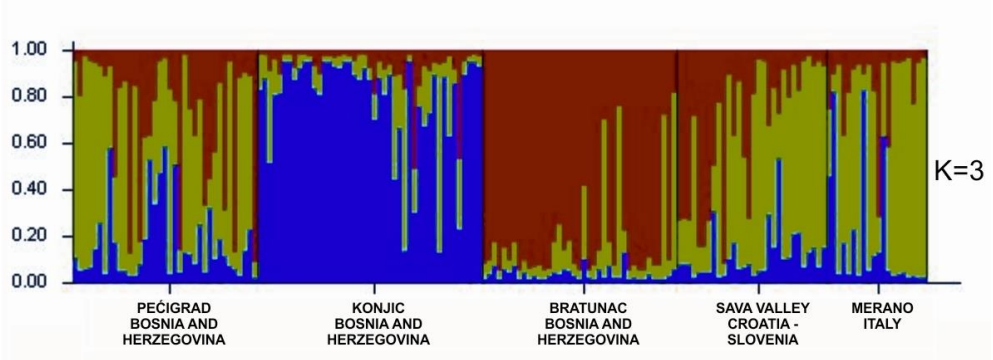


Figure 3. Population structure of 171 *Castanea sativa* samples, assigned to three clusters using Bayesian model-based cluster procedure within Structure version 2.2.3 (Pritchard et al. 2000).

*Table 5. Analysis of molecular variance (AMOVA) based on 5 SSR loci for Structure clusters (K=3 defined by Structure (Pritchard et al. 2000) (genotypes with qI>75%).*

| Source of variation | df | Variance components | Total variance (%) | $f_{CT}$ | P-value |
|---|---|---|---|---|---|
| Structure cluster (K=3) | | | | | |
| Within groups: | 229 | 0.77863 Vb | 22.28 | 0.223 | <0.01 |
| Among groups: | 2 | 0.22318 Va | 77.72 | | |
| Total | 231 | 1.00181 | | | |

d.f. degree of freedom.

## DISCUSSION

This is the first known large scale study on chestnut which includes analyses of more than ten different genomic microsatellites. The average number of alleles per locus (5.95), obtained in our study, is lower than the values reported by BECCARO *et al*. (2012) (7.4) and MARTÍN *et al*. (2012) (13.14), who also examined wild chestnut populations. The mentioned studies also published higher mean gene diversity (0.676 and 0.804), than the value reported here (0.605). However, considering that both BECCARO *et al*. (2012) and MARTÍN *et al*. (2012) analyzed a smaller number of highly polymorphic SSR loci (nine and six microsatellites, respectively), this is not unexpected.

The main criteria for the selection of microsatellites in previous chestnut studies has been the polymorphism of the loci, but also the amount of coverage of chestnut linkage groups (PEREIRA-LORENZO *et al*., 2010; BECCARO *et al*., 2012; MARTÍN *et al*., 2012; MELLANO *et al*., 2012; MATTIONI *et al*., 2013). Although, all key studies cited here have successfully selected and analyzed microsatellites that are highly polymorphic, physically unlinked and covering most of the chestnut chromosomes, there is a drawback to using a low number of SSR loci in diversity studies. Namely, PEREIRA-LORENZO *et al*. (2010) who analyzed 787 chestnut accessions using 10 microsatellites, corresponding to 11 loci, had to remove three loci and one population due to their deviation from HWE, caused by the presence of null alleles. MATTIONI *et al*. (2013) who analyzed 779 wild chestnut trees, sampled from 31 European sites, using six SSRs, had to discard EMCS25 microsatellite due to high frequencies of null alleles. Even though a correction can be applied for the presence of null alleles, as demonstrated in our study, the number of available statistical approaches requiring as an input corrected allele frequencies and not genotype data is severely restricted. Therefore, the results of SSR polymorphism presented here (Table 2) can serve as a valuable information for future molecular studies on chestnut, which seek additional markers to replace ones with high null allele frequency.

Molecular data, obtained in our study, revealed much about the genetic structure of the analyzed populations. Considering the clear differentiation of southern and eastern B&H chestnut populations (Konjic and Bratunac) from the Italian population, determined by pairwise Gst, FCA and Bayesian Structure analyses, it is highly probable that these populations originated from independent shelter zones (refugia), after the last glaciation period. Although fossil pollen data indicates a high likelihood of *Castanea* refugia in the Balkan Peninsula (KREBS *et al*., 2004), the latest major molecular study on chestnut, conducted by MATTIONI *et al.* (2013), proposes that chestnut populations found in the southern end of the Balkans (Greece) originated from a western Turkey refugium.

Introgression of fruit germplasm, from Turkey to western Balkan, has previously been reported by BURAK *et al*. (2014) and AKÇAY *et al*. (2014), who analyzed the SSR data obtained from traditional Turkish apple and pear cultivars respectively, together with the microsatellite profiles published by GASI *et al*. (2010) and GASI *et al*. (2013), on traditional apple and pear genotypes from Bosnia and Herzegovina.

However, since MATTIONI *et al*. (2013) study did not include samples from large parts of the Balkans, further investigation are needed in order to determine the possible existence and location of chestnut shelter zones in this part of Europe.

Strong genetic differentiation registered between southern and eastern B&H chestnut populations (Konjic and Bratunac) can be explained by genetic adaptations to notably different climatic conditions present in the south (Mediterranean climate) and east (Continental climate) of Bosnia and Herzegovina. Genetic adaptation to climatic condition has previously been reported by PEREIRA-LORENZO *et al.* (2010) as the probable cause for the differentiation between chestnut accessions from northern and central Iberian Peninsula. However, the fact that the strongest genetic differentiation among all analyzed populations was detected between Konjic and Bratunac could also indicate that current populations in western Balkans have originated from different shelter zones, with distinct climatic conditions. In fact a similar situation was reported by MATTIONI *et al*. (2013) on two chestnut gene pools in Turkey, which originated from different glacial refugia and remained separated after the glacial period, due to the large climatic differences between eastern and western Turkey. LUSINI *et al*. (2014) likewise reported the presence of a genetically divergent and homogenous chestnut population in Bulgaria (Slavyanka population), presumably a result of an ancient recolonization from a distinct refugium, followed by a subsequent genetic isolation.

Unlike the chestnut samples from the southern and eastern Bosnia and Herzegovina, individuals from Pećigrad (northwest B&H) and Sava valley (Slovenia-Croatia) populations belonged to two different clusters (gene pools), defined by Bayesian structure analyses. Based on our results we propose a presence of an introgression zone in the northwestern Balkans, established through gene flow from the Italian and the Balkan Peninsula. Conclusion of MATTIONI *et al*. (2013) that the contributing factor for the existence of the same chestnut gene pool in Italian and Iberian Peninsulas is due to human mediated transport of the plant material, can also be applied here. Geographical proximity of the Italian, Slovenian-Croatian and northwestern B&H chestnut populations (Figure 1), analyzed in this study, further supports the likelihood of an introgression zone.

We believe that the results of our study contribute to the overall knowledge about the expansion process of *C. sativa* during the postglacial period. Analyses of population structure revealed the presence of a highly differentiated chestnut population in southern Bosnia and Herzegovina, and two admixed populations in northwestern Balkans. Obtained molecular data will help form the bases for the overall assessment of this genetic resource. The results of this study will significantly contribute towards establishing a regional conservation and utilization strategy, in regards to European chestnut in western the Balkans.

ACKNOWLEDGEMENT

This study was partially funded by the Norwegian government through HERD (Program for Higher Education, Research and Development) project “Evaluation of fruit genetic resources in Bosnia-Herzegovina with the aim of sustainable, commercial utilization” ref. no. 332160 UE.

The authors thank Professor Clive Kaiser for his critical review of the manuscript.

Received June 13th, 2016
Accepted March 25th, 2017

REFERENCES

AKÇAY, M.E., M. BURAK, K. KAZAN, C. YUKSEL, F. MUTAF, M. BAKIR, H. AYANOGLU, A. ERGÜL (2014): Genetic analysis of Anatolian pear germplasm by simple sequence repeats (SSRs). Ann. Appl. Biol., *164***:** 441–452.

BARRENECHE, T., M. CASASOLI, K. RUSSEL, A. AKKAK, H. MEDDOUR, C. PLOMION (2004): Comparative mapping between *Quercus* and *Castanea* using simple-sequence repeats (SSRs). Theor. Appl. Genet., *108*: 558–566.

BECCARO, G.L., D.T. MARINONI, G. BINELLI, D. DONNO, P. BOCCACCI, R. BOTTA, A.K. CERUTTI, M. CONEDERA (2012): Insights in the chestnut genetic diversity in Canton Ticino (Southern Switzerland). Silvae Genet., *61*: 6.

BELKHIR, K., P. BORSA, L. CHIKHI, N. RAUFAST, F. BONHOMME (2001): GENETIX 4.02, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

BUCK, E.J., M. HADONOU, C.J. JAMES, D. BLAKESLEY, K. RUSSEL (2003): Isolation and characterization of polymorphic microsatellites in European chestnut (*Castanea sativa* Mill.). Mol. Ecol. Notes, *3*: 239–241.

BURAK, M, A. ERGULA, K. KAZAN, M. AKCAY, C. YUKSEL, M. BAKIR, F. MUSTAF, E. AKPINAR E, S. YASASIN, H. AYANOGLU (2014): Genetic analysis of Anatolian apples (*Malus* sp.) by simple sequence repeats. J. Syst. Evol., *52*: 580-588.

CASASOLI, M., J. DERORY, C. MORERA-DUTREY, O. BRENDEL, I. PORTH, J.M. GUEHL, F. VILLANI, A. KREMER (2006): Comparison of quantitative trait loci for adaptive traits between oak and chestnut based on an expressed sequence tag consensus map. Genetics, *172*: 533–546.

CHAKRABORTY, R., M. KIMMEL, D.N. STIVERS, L.J. DAVISON, R. DEKA (1997): Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc. Natl. Acad. Sci. USA, *94*: 1041–1046.

CHAPUIS, M.P., A. ESTOUP (2007): Microsatellite null alleles and estimation of population differentiation. Mol. Biol. Evol., *24*: 621–631.

CONEDERA, M., P. KREBS, W. TINNER, M. PRANDELLA, D. TORRIANI (2004): The cultivation of *Castanea sativa* (Mill.) in Europe, from its origin to its diffusion on a continental scale. Veg. Hist. Archaeobot., *13*: 161–179.

DOYLE, J.J., J.L. DOYLE (1990): Isolation of plant DNA from fresh tissue. Focus, *12*: 13-15.

EARL, D.A., B.M. VON HOLDT (2011): Structure harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour., *4*: 359–361.

EVANNO, G., S. REGNAUT, J. GOUDET (2005): Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol., *14*: 2611–2620.

EXCOFFIER, L., H.E.L. LISCHER (2010): Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour., *10*: 564–567.

EXOFFIER, L., P.E. SMOUSE, J.M. QUATTRO (1992): Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics, *131*: 479–491.

GASI, F., M. KURTOVIC, B. KALAMUJIC, N. POJSKIC, J. GRAHIC, C. KAISER, M. MELAND (2013): Assessment of European pear (*Pyrus communis* L.) genetic resources in Bosnia and Herzegovina using microsatellite markers. Sci. Hort., *157*: 74–83.

GASI, F., S. SIMON, N. POJSKIC, M. KURTOVIC, I. PEJIC (2010): Genetic assessment of apple germplasm in Bosnia and Herzegovina using microsatellite and morphologic markers. Sci. Hort, *126*: 164-171.

GOUDET, J. (2002): FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2). http://www.unil.ch/izea/softwares/fstat.html.

GUO, S.W., E.A. THOMPSON (1992): Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics, *48*: 361-372.

HAMMER, O., D.A.T. HARPER, P.D. RYAN (2001): Past: Paleontological Statistics Software Package for education and data analysis. Palaeontol. Electron., *4*(1): art. 4.

IDŽOJTIĆ, M., M., ZEBEC, I., POLJAK, J., MEDAK, (2009): Variation of sweet chestnut (*Castanea sativa* Mill.) populations in Croatia according to the morphology of fruits. Sauteria, *18*: 323-333.

KALINOWSKI, S.T., M.L. TAPER (2006): Maximum likelihood estimation of the frequency of null alleles at microsatellite loci. Conserv. Genet., *7*: 991-995.

KREBS, P., M. CONEDERA, M. PRADELLA, D. TORRIANI, M. FELBER, W. TINNER (2004): Quaternary refugia of the sweet chestnut (*Castanea sativa* Mill.): an extended palynological approach. Veg. Hist. Archaeobot., *13*: 145–160.

LUSINI, I., I. VELICHKOV, P. POLLEGIONI, F. CHIOCCHINI, G. HINKOV, T. ZLATANOV, M. CHERUBINI, C. MATTIONI (2014): Estimating the genetic diversity and spatial structure of Bulgarian *Castanea sativa* populations by SSRs: implications for conservation. Conserv. Genet., *15*: 283-293.

MANTEL, N. (1967): The detection of disease clustering and a generalized regression approach. Cancer Res., *27*: 209-220.

MARINONI, D.T., A. AKKAK, C. BELTRAMO, P. GUARALDO, P. BOCCACCI, G. BOUNOUS, A.M. FERRARA, A. EBONE, E. VIOTTO, R. BOTTA (2013): Genetic and morphological characterization of chestnut (*Castanea sativa* Mill.) germplasm in Piedmont (north-western Italy). Tree Genet. Genomes, *9*: 1017–1030.

MARINONI, D., A. AKKAK, G. BOUNOUS, K.J. EDWARDS, R. BOTTA (2003): Development and characterization of microsatellite markers in *Castanea sativa* (Mill.). Mol. Breeding., *11*: 127–136.

MARTÍN, M.A., C. MATTIONI, J.R. MOLINA, J.B. ALVAREZ, M. CHERUBINI, M.A. HERRERA, F. VILLANI, L.M. MARTÍN (2012): Landscape genetic structure of chestnut (*Castanea sativa* Mill.) in Spain. Tree Genet. Genomes, *8*: 127–136.

MARTÍN, M.A., C. MATTIONI, M. CHERUBINI, D. TAURCHINI, F. VILLANI (2010): Genetics diversity in European chestnut populations by means of genomic and genic microsatellite markers. Tree Genet. Genomes, *6*: 735–744.

MATTIONI, C., M.A. MARTÍN, P. POLLEGIONI, M. CHERUBINI, F. VILLANI (2013): Microsatellite markers reveal a strong geographical structure in European populations of *Castanea sativa* (*Fagaceae*): evidence for multiple glacial refugia. Am. J. Bot., *100*: 1-11.

MELLANO, M.G., G.L. BECCARO, D. DONNO, D.T. MARINONI, P. BOCCACCI, S. CANTERINO, A.K. CERUTTI, G. BOUNOUS (2012): *Castanea* spp. biodiversity conservation: collection and characterization of the genetic diversity of an endangered species. Genet. Resour. Crop Ev., *59*: 1727-1741.

NEI, M. (1987): Molecular evolutionary genetics. New York: Columbia University Press.

ORDONEZ, A., SVENNING, J.-C. (2017): Consistent role of Quaternary climate change in shaping current plant functional diversity patterns across European plant orders. Sci. Rep., *7*: 42988.

OTA, T. (1993): DISPAN: Genetic distance and phylogenetic analysis. University Park, Pennsylvania State University, http://evolution.genetics.washington.edu/phylip/software.dist.html.

PEAKALL, R., P.E. SMOUSE (2012): GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics, *28*: 2537-2539.

PEREIRA-LORENZO, S., R.M.L. COSTA, A.M. RAMOS-CABRER, C.A.M. RIBEIRO, M.F.S. DA SILVA, G. MANZANO, T. BARRENECHE (2010): Variation in grafted European chestnut and hybrids by microsatellites reveals two main origins in the Iberian Peninsula. Tree Genet. Genomes, *6*: 701-715.

PRITCHARD, J.K., M. STEPHENS, P. DONNELLY (2000): Inference of population structure using multilocus genotype data. Genetics, *155*: 945–959.

VAN OOSTERHOUT, C., W.F. HUTCHINSON, D. WILLS, P. SHIPLEY (2004): MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes, *4*: 535-538.

# GENETIČKA STRUKTURA I DIVERZITET POPULACIJA EVROPSKOG KESTENA (*Castanea sativa* MILL.) NA ZAPADNOM BALKANU: NA RASKRSNICI IZMEĐU ISTOKA I ZAPADA

Azra SKENDER[1], Mirsad KURTOVIĆ[2], Naris POJSKIĆ[3], Belma KALAMUJIĆ STROIL[3*], Semina HADŽIABULIĆ[4], Fuad GAŠI[2]

[1]Biotehnički fakultet, Univerzitet u Bihaću
[2]Poljoprivredno-prehrambeni fakultet, Univerzitet u Sarajevu
[3]Institut za genetičko inženjerstvo i biotehnologiju, Laboratorija za molekularnu genetiku prirodnih resursa, Univerzitet u Sarajevu
[4]Agromediteranski fakultet, Univerzitet Džemal Bijedić u Mostaru

Izvod

Evropski kesten (*Castanea sativa* Mill.) veoma je cijenjen kao izvor plodova i drveta, ali i kao važan izvor nektara i polena u proizvodnji meda na zapadnom Balkanu. U ovoj studiji analizirane su, primjenom 21 mikrosatelitnog markera, četiri populacije kestena sa područja zapadnog Balkana i referentna populacija iz sjeverne Italije. Najveći genetički diverzitet uočen je u populacijama geografski najbližim italijanskom poluotoku, koje su ujedno pokazale i najveći stepen sličnosti s uzorcima iz Italije. Najveća genetička diferencijacija primijećena je između sjevernih i istočnih populacija kestena iz Bosne i Hercegovine (B&H) (Gst=12,05%). Ova izražena diferencijacija vjerovatno je uslovljena genetičkim adaptacijama na značajno različite klimatske uslove prisutne na jugu (mediteranska klima) i istoku (kontinentalna klima) Bosne i Hercegovine. Jasna genetička diferencijacija južnih i istočnih bh. populacija kestena od italijanskih populacija, uočena na osnovu Gst, FCA i STRUCTURE analize, ukazuje na to da su ove populacije najvjerovatnije potekle iz nezavisnih refugijuma nakon posljednjeg glacijalnog perioda. Na osnovu ovih rezultata, smatramo da postoji introgresijska zona na sjeverozapadu Balkana, uspostavljena genskim protokom između Apeninskog i Balkanskog poluotoka. Dobivena saznanja o strukturi analiziranih populacija značajno će doprinijeti uspostavljanju regionalnih strategija za konzervaciju i upravljanje populacijama evropskog kestena na zapadnom Balkanu.

Primljeno 13.VI.2016.
Odobreno 25.III. 2017.